ONTOLOGICAL APPROACH FOR IMPROVING SEMANTIC WEB SEARCH RESULTS

Awes Siddiqui [1], P. A. Tijare[2]

[1,2]*Computer Science, Sipna COET, Amravati, India, awes.siddiqui@gmail.com, pritishtijare@rediffmail.com*

Abstract

we propose personalized information system to provide more user-oriented information considering context information such as a personal profile based, location based and click based. Our system can provide associated search results from relations between the objects using context ontologies modeling created by the categorized layers data. Based on the similarities between item descriptions and user profiles, and the semantic relations between concepts, content based and collaborative recommendation models are supported by the system. User defined rules based on the ontology description of the service and interoperates within any service domain that has an ontology description.

Keywords- Ontology; Semantic Web; RDF;OWL

--******--

1. INTRODUCTION

The Semantic Web is an extension of the current Web [1] that allows the meaning of information to be precisely described in terms of well-defined vocabularies that are understood by people and computers. On the Semantic Web information is described using a new W3C standard called the Resource Description Framework (RDF). Semantic Web Search is a search engine for the Semantic Web. Current Web sites can be used by both people and computers to precisely locate and gather information published on the Semantic Web. Ontology [2] is one of the most important concepts used in the semantic web infrastructure, and Resource Description Framework/Schema RDF(S) and Web Ontology Languages (OWL) are two W3C recommended data representation models which are used to represent ontologies. The Semantic Web will support more efficient discovery, automation, integration and reuse of data and provide support for interoperability problem which cannot be resolved with current web technologies. Currently research on semantic web search engines are in the beginning stage, as the traditional search engines such as Google, Yahoo, and Bing (MSN) and so forth still dominate the present markets of search engines.

Web Search engines do an excellent job when the queries are understandable and exact. Generally, user queries are short, ambiguous and not well-formed in nature. Ambiguous queries confuse the search engine and not satisfy the specific needs of the user. Search engines should provide precise search results to the end user. When queries are issued to the search engines they return the same results to user queries irrespective of topical interest or context. Different users may send query to the search engines that are short and ambiguous. Sometimes the same query may search for different information needs and

purposes. But the system will never be able to provide users' precise needs but it provides in general.

Personalization of search engine is not effective on some queries. Search engine respond to the list of ranked pages based on the relevance of the query. So that search engines generate user profiles to identify and get the users' actual needs.

World Wide Web has been advancing towards greater personalization. Services on the Web such as, social networking, e-commerce or search sites, store user information in order to profile the user and target specific products or ads of interest. Since web service functionality is increasingly relying on user information, a user's context is becoming more crucial towards creating a personalized set of services within the Web.

2. RELATED WORK

In general, context-aware systems are constructed to be adaptive to user needs according to her/his profile and context. For instance, we mention Location-Based Services (LBS) as a general example about geographic location and context [3]. Ontology technology [4] is considered as a solution to integrate and to analyze dispersed data sources from vender offerings and Web content. Some researches [5] have suggested advantages by expressing concepts through ontology in context aware systems. Moreover, ontology-based services [6] can provide augmented, enriched, and synthesized information using suitable reasoning mechanisms and more precise and adaptive information for a user's current context conditions by integrating, sharing, and exchanging items. However, the search services based on the ontology structure have limitations in implementing real-world applications

because of the shortage of existing content as compared to databases. To allow large-scale assessment of personalized search, an evaluation framework was developed on query logs and this approach has improved search precision on few selected queries but damage other queries and far from finest search. Current researches of context-aware systems using context information have been processed in the fields such as Web-based context-aware architecture, context aware LBS application, and ontology-based context modeling. Present is a service-oriented context aware system based on the LBS, which provides several services such as map presentation, routing and advertisement services. The context data of the system is gathered by connecting other Web services. Also, the system expressed user's context information using ontology.

The defined ontology is categorized into user, profile, task, appointment, physiologic, Emotion class, and each class contains some other sub-classes. However, the system is lacking in detail personalized search information for layers such as hotels, pools, shops in supporting map services, only presenting Personal of interest of user interest provides a software architecture which can be used for different types of context systems to provide context information according to user needs.

3. ONTOLOGIES IN SEMANTIC WEB

A. Location Ontology (Heading 2)

If a keyword/phrase exists frequently in the web-snippets arising from the query q, it represents an important concept related to the query, as it coexists in close proximity with the query in the top documents. Thus, our content concept extraction method first extracts all the keywords and phrases from the web-snippets arising from q. After obtaining a set of keywords/phrases (ci), the following support formula, which is inspired by the well-known problem of finding frequent item sets in data mining, is employed to measure the interestingness of a particular keyword/phrase ci with respect to the query q:

$$support\ (c_i) = sf\ (c_i)/n\ .|c_i|$$

where $sf\ (c_i)$ is the snippet frequency of the keyword/phrase c_i (i.e. the number of web-snippets containing c_i), n is the number of web-snippets returned and $|c_i|$ is the number of terms in the keyword/phrase c_i. If the support of a keyword/phrase c_i is higher than the threshold s (s = 0:03 in our experiments), where c_i is a concept for the query q.

As mentioned, the ontologies are used to maintain concepts and their relationships extracted from search results. The location ontology is built here to represent these content concepts. The location ontology is built based on the following types of relationships for content concepts.

Similarity:
Two concepts which coexist a lot on the search results might represent the same topical interest. If coexist (ci, cj) > _1 (_1 is a threshold), then ci and cj are considered as similar.

Parent-Child Relationship:
More specific concepts often appear with general terms, while the reverse is not true. Thus, if pr (cj, ci) > _2 (_2 is a threshold), where ci as cj's child. Fig 1 shows an example content ontology created for the query 'apple'. Content concepts linked with a double sided arrow ($) are similar concepts, while concepts linked with a one-sided arrow (!) are parent-child concepts. The ontology shows the possible concept space arising from a user's queries.
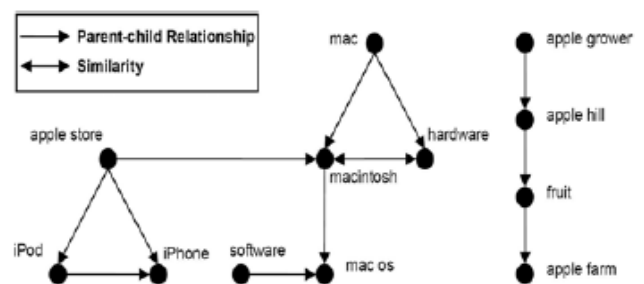

Fig. Apple Example

B. Content Ontology

The approach for extracting location concepts is different from that for extracting content concepts. First, a web-snippet usually embodies only a few location concepts. As a result, very few of them co-occur with the query terms in web snippets. To alleviate this problem, the location concepts are extracted from the full documents. The content ontology is built to represent these location concepts.

Second, due to the small number of location concepts embodied in documents, the similarity and parent-child relationship cannot be accurately derived statistically. Additionally, the content ontology extraction method extracts all of the keywords and key-phrases from the documents returned for q. If a keyword or key-phrase in a retrieved document matches a location name in the predefined location ontology, it will be treated as a location concept of d. Similar to the content ontology, locations are assigned with different weights according the user's click through.

C. Pattern Based Ontology

Reasoning:
Some features of ontologies cause performance problems for certain reasoners. Having information about these features, possibly gathered via machine learning methods, we can transform parts of ontologies with such problematic entities.

Modularization:

 Modular ontologies are a prerequisite for effective knowledge sharing. However, if the source and target ontology are modeled using different styles (such as property- vs. relation-centric), the user faces difficulties when choosing fragments to be imported. Modeling style of the various modules.

Matching:

Most Ontology Matching (OM) tools deliver simple entity-to-entity correspondences. When applying lexical methods on names of entities, they ignore matching of complex structures. Complex matching can be mediated by correspondence patterns, which however most OM tools do not support. Attempting to transform, prior to matching, ontology to its variant using transformation patterns, could thus help the OM tools.

4. ANALYSIS OF PROBLEM

The existing profile-based [7] personalized search approaches are not consistent when compared to click-based method. To allow large-scale assessment of personalized search, an evaluation framework was developed on query logs.

It is difficult for clients to discover more appropriate information for their search query. This is because of increase of maximum users of internet and amount of web pages. It takes time to search results for users' particular needs. To get relevant information users should go for a public search engine and need to submit their query. But this is also rendering most irrelevant results. So users may be confused with the results and the problem arises here because of not providing users' actual needs in a well formed structure. Previous methods are not accurate in capturing user interest and profiling is not reliable.

The differences between our work and existing works are:
Existing works require the users' to manually define their location preferences explicitly (with latitude-longitude pairs or text form). With the automatically generated content and location user profiles, our method does not require users to explicitly define their location interest manually.

The conceptual relationship between the documents has to be represented in order to identify the information that a user wants from those represented concepts. To represent the semantic relation, the ontology is used here. To build a user profile [8], the Web pages that the user visited are monitored and the system represents the long-term and short-term preference weights as the preference ontology after inferring relevant concepts from the general ontology. At the recommendation stage, the system recommends documents according to user preference concepts and document similarity measure.

5. PROBLEM DEFINITION

- **Low precision and high recall**

Some Intelligent semantic search engines cannot show their significant performance in improving precision and lowering recall. In Ding's semantic flash search engine, the resource of the search engine is based on the top-50 returned results from Google that is not a semantic search engine, which could be low precision and high recall [9].

- **Identity intention of the user**

User intention identification plays an important role in the intelligent semantic search engine. For example, in chiung-Hon leon lee introduced a method for analyzing the request terms to fit user intention, so that the service provided will be more suitable for the user [10].

- **Individual user patterns can be extrapolated to global users.**

In early search engine that offered disambiguation to search terms. A user could enter in a search term that was ambiguous (e.g., Java) and the search engine would return a list of alternatives (coffee, programming language, island in the South Seas).

- **Inaccurate queries.**

We have user typically domain specific knowledge. And users don't include all potential Synonyms and variations in the query, actually user have a problem but aren't sure how to phrase.

6. PROPOSED WORK

The user's profile is stored as a Resource Description Framework (RDF). Content format based on the defined ontology. Next, users attain results of associated search based on Google the ontology and database as to the following process.
(1) The Google search result gives the semantic data after queries ranking.
(2) The user distance from current location on the configuration menu, the system operates searches with the inputted values and the profile with the user information.
(3) The search is executed that the detail information is queried from the database through the object identifier of geospatial data after the user's preference codes is extracted from the defined service type ontology.
(4) The click based pattern searching on of context-based personalized search.

It is important to extract real and dynamic context information through Web services and sensors and to provide adaptive information to user's preferences. To do this, we need to implement an available Context aware system in the real-world by gathering more realistic context information. Also, this system will include an ontology-based search service with inference function considering the user's history and profile.

The personalization approach is based on concepts to profile the interests and preferences of a user. An issue to be addressed is how to extract and represent concepts from search results of the user. An OMF profiling method is proposed in which concepts can be further classified into different types, such as content concepts (location ontology), location concepts (content ontology), name entities, dates etc.

Google Search

First Step involves of query formation with these collection of words. Permutations and combinations are needed to form various refined queries from the words obtained. The queries formed will be more refined and will fetch more semantically related web links on passing these queries as input to the search engines. The refined queries are sent to search API which fetches the web links related to the user query. This can be accomplished with the help of **Google Search API**.

Location Based Ranking Data

The web links obtained after passing the refined queries to the search API are now filtered and ranked to make it more refined. If any of the web links are not relevant to the given query they are filtered out. Ranking is applied to all web links obtained from all possible queries formed under permutation. On applying ranking the web links are re-ranked in the appropriate order of semantic relatedness. This can be accomplished with the help of **Ranking Algorithm**.

Profile Based Ranking

This process consists of individual user's interests, topical categories of user interests and identifies the relationship among the concepts. The proposed approach was based on topic ontology for concept based user profile generation from search engine logs. Spreading activation algorithm was used to optimize the relevance of search engine results. Topic ontology was constructed to identify the user interest by assigning activation values and explore the topics similarity of user preferences. This can be accomplished with the help of User Profile Checking.

Click Through Based Ranking

The meta search engine collects click through data from the users and performs personalized ranking of the search results based on the learnt profiles of the users. The users are invited to submit totally test queries to our metasearch engine. For each query submitted, the top search results are returned to the users. The topical categories of the test queries. This can be accomplished with the help of **Pattern Based Ranking**.

CONCLUSIONS

Personalization framework is proposed for automatically extracting and learning a user's content and location preferences based on the user's click through. Different methods are developed for extracting content and location concepts, which are maintained along with their relationships in the content and location ontologies. The notion of content

and location entropies is introduced to measure the diversity of content and location information associated with a query and click content and location entropies to capture the breadth of the user's interests in these two types of information.

REFERENCES

[1]   Capilla, R. "Context-aware Architectures for Building Service-Oriented Systems". Proceeding of the Conference on Software Maintenance and Reengineering, IEEE, pp300-303.

[2]   Berners-Lee, T., Hendler, J., Lassila, O. "The Semantic Web". Scientific American, Vol. 284. No. 5, pp34-43.

[3]   Uschold, M., Gruning, M. "Ontologies Principles, Methods, and Application". Artificial Intelligence Applications Institute. The University of Edinburgh.

[4]   Devedzic, V. "Understanding Ontological Engineering". Communications of the ACM. Vol. 45 No. 4. Pp136-144.

[5]   Prabaharan, S. and 2R.S.D. "Ontological Approach for Effective Generation of Concept Based User Profiles to Personalize Search Results", Wahidabanu Journal of Computer Science 8 (2): 205-215, ISSN 1549-3636 Science Publications.

[6]   G.Madhu and Dr.A.Govardhan Dr.T.V.Rajinikanth. "Intelligent Semantic Web Search Engines: A Brief Survey", International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1.

[7]   Yousif Al Ridhawi, Ismaeel Al Ridhawi, Loubet Bruno, and Ahmed Karmouch. "Policy-Based Personalized Context Dissemination for Location-Aware Services" P. Sénac, M. Ott, and A. Seneviratne (Eds.): MobiQuitous.

[8]   Rajasurya, mudhuMuralidharan, Sandhiya Devi, Dr.S.Swamynathan. "Semantic Information Retrieval Using Ontology in University Domain".

[9]   G.Monica K.Manoharan. "Semantic Search Engine Using Content and Location   Ontology Extraction" International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555, Vol. 2, No. 1.Ondrej Svab-Zamazal, Vojtech Svatek. "Pattern-Based Ontology Transformation Service", In Znalosti.